Application of Bayesian methods to modelling of extreme precipitation events

Kári Hlynsson

University of Iceland

January 2025

Introduction (1/2)

- Oct. 29, 2024: $\approx$ 400 mm of rain ($\approx$ 1 year's worth) fell in 8 hours across Valencia, Castilla-La Mancha, and Andalusia, Spain.
- Casualties: 232 dead, 3 missing.
- Damages: $\approx$ €1 billion to housing and infrastructure.
- Modelling *extreme precipitation events* vital in mitigating damage to infrastructure and human life, and elucidating links to warming climate

Introduction (2/2)

Fisher - Tippett theorem.
If $(X_n)_{n\in\mathbf{N}}$ are i.i.d., $M_n := \max\{X_1, \ldots, X_n\}$, and there exist normalising sequences $a_n \in \mathbf{R}_+$ and $b_n \in \mathbf{R}$ such that

$$\lim_{n\to\infty} \Pr\left\{\frac{M_n - b_n}{a_n} \leq y\right\} \to G(y),$$

then necessarily $G(y) = \mathrm{GEV}(y \mid \mu, \sigma, \xi)$, where

$$\mathrm{GEV}(y \mid \mu, \sigma, \xi) = \begin{cases} \exp\left(-\left[1 + \xi(y-\mu)/\sigma\right]_+^{-1/\xi}\right), & \xi \neq 0, \\ \exp\left(-\exp\left[-(y-\mu)/\sigma\right]\right), & \xi = 0. \end{cases}$$

Overview

- Bayesian modelling framework
- Model selection and evaluation
- Results and discussion

The data (1/2)

- Raw data from Southern England spans ≈ 132 years of daily cumulative precipitation measurements in millimetres over 24 hour period.

- Aggregated into annual block maxima

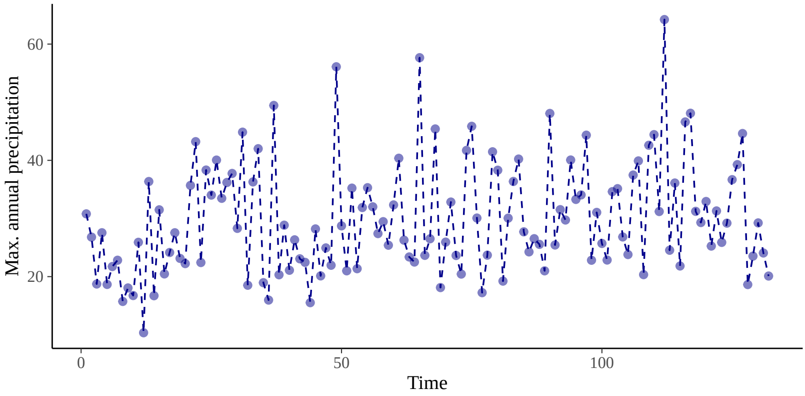
$$y_t = \max\{X_1^{(t)}, \ldots, X_{365}^{(t)}\},$$

  where $X_1^{(t)}, \ldots, X_{365}^{(t)}$ are daily measurements from year t.

- Result is a data set numbering $n = 132$ measurements.

- Interested in modelling *extreme values* in $y = (y_1, \ldots, y_n)$.

The data (2/2)

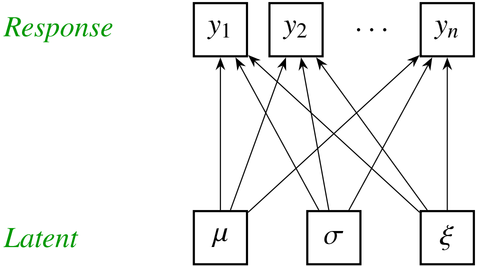


Modelling framework

- Response layer y is assigned generalised extreme value (GEV) distribution
- Compare *stationary* models with fixed parameters and *non-stationary* models with evolving location parameter over time, μ_t. Does a time trend improve model fit?
- MCMC applied to sample posterior $\pi(\theta \mid y)$. 1000 samples for burn-in, 8000 for sampling. Convergence estimated by trace plots, ACF, and $\hat{R}$.
- LOOIC / WAIC and PPC for model selection and evaluation

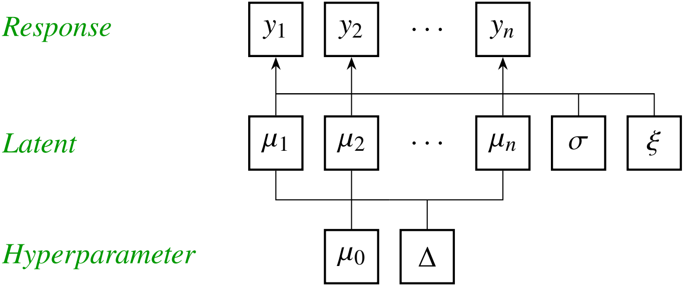
Stationary model

- Response layer: $y \sim \text{GEV}(\mu, \sigma, \xi)$
- Latent layer: $\mu \sim \text{N}(0, \sigma_\mu)$, $\sigma \sim \text{Exp}(\lambda_\sigma)$, $\xi \sim \text{Beta}(\alpha_\xi, \beta_\xi)$ on interval $(-0.5, 0.5)$.
- $\sigma_\mu = 10$, $\lambda_\sigma = 3$, $\alpha_\xi = \beta_\xi = 4$



Non-stationary model (linear time trend)

- Response layer: $y_t \sim \mathrm{GEV}(\mu_t, \sigma, \xi)$
- Latent layer: $\mu_t = \mu_0(1 + \Delta(t - t_c))$, $\sigma \sim \mathrm{Exp}(\lambda_\sigma)$, $\xi \sim \mathrm{Beta}(\alpha_\xi, \beta_\xi)$ on $(-0.5, 0.5)$
- Hyperparameter layer: $\mu_0 \sim \mathrm{N}(0, \sigma_{\mu_0})$, $\Delta \sim \mathrm{N}(0, \sigma_\Delta)$, where $\sigma_{\mu_0} = \sigma_\Delta = 10$, $t_c = \lfloor N/2 \rfloor$

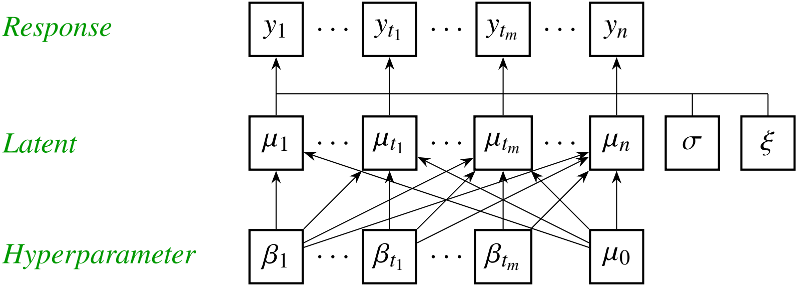


Non-stationary model (piecewise time trend) (1/3)

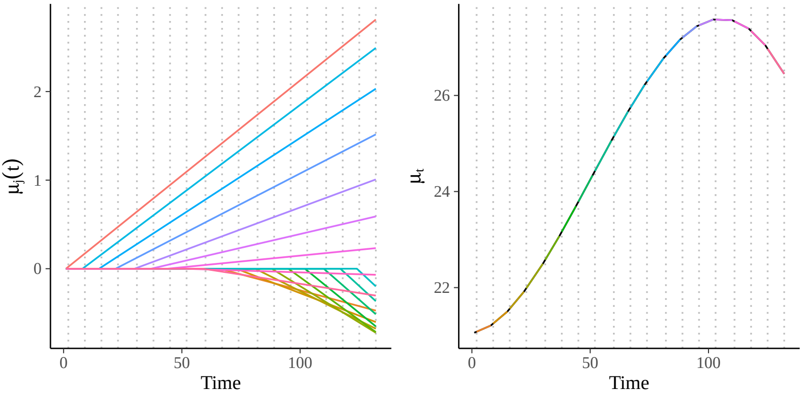
- Partition $T = \{1, \ldots, n\}$ into m equally spaced time points, $1 = t_1 < \cdots < t_m < n$. The partition granularity m is tuned for optimal results.
- Response layer: $y_t \sim \mathrm{GEV}(\mu_t, \sigma, \xi)$
- Latent layer: $\mu_t = \mu_0 + \sum_{j=1}^{m} \mu_j(t)$ where $\mu_j(t) = \beta_j \mathbf{1}_{\{t \geq t_j\}}(t - t_j)$.
- Hyperparameter layer: $\mu_0 \sim \mathrm{N}(0, \sigma_{\mu_0})$, $\beta_1 \sim \mathrm{N}(0, \sigma_\beta^{(1)})$ with $\sigma_{\mu_0} = 10$, and $\sigma_\beta^{(1)} = 0.10$. Three priors for remaining β_j:
  - I.I.D.: $\beta_{j+1} \sim \mathrm{N}(0, \sigma_\beta)$ for $j \geq 1$,
  - R.W.: $\beta_{j+1} \sim \mathrm{N}(\beta_j, \sigma_\beta)$ for $j \geq 1$,
  - AR(1): $\beta_{j+1} \sim \mathrm{N}(\phi\beta_j, \sigma_\beta)$ for $j \geq 1$, $\phi \sim \mathrm{Unf}(-1, 1)$,

  where $\sigma_\beta = 0.01$.

Non-stationary model (piecewise time trend) (2/3)



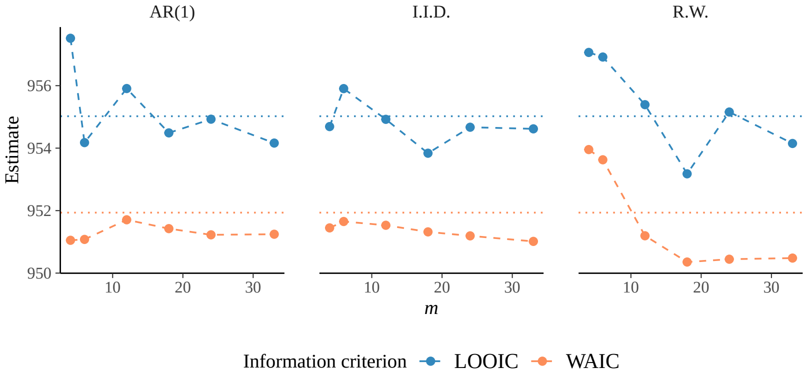
Non-stationary model (piecewise time trend) (3/3)



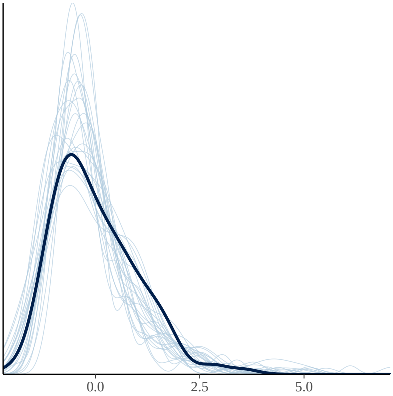
Model selection (1/2)

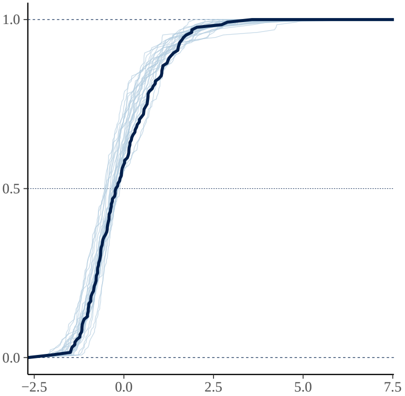
	Stationary	Linear trend	Piecewise trend		
			AR(1)	I.I.D.	R.W.
$\widehat{\text{LOOIC}}$	965.2	955.0	954.2 (6)	953.8 (18)	**953.2 (18)**
$\widehat{\text{WAIC}}$	961.9	951.9	951.1 (4)	951.0 (33)	**950.3 (18)**

Model selection (2/2)

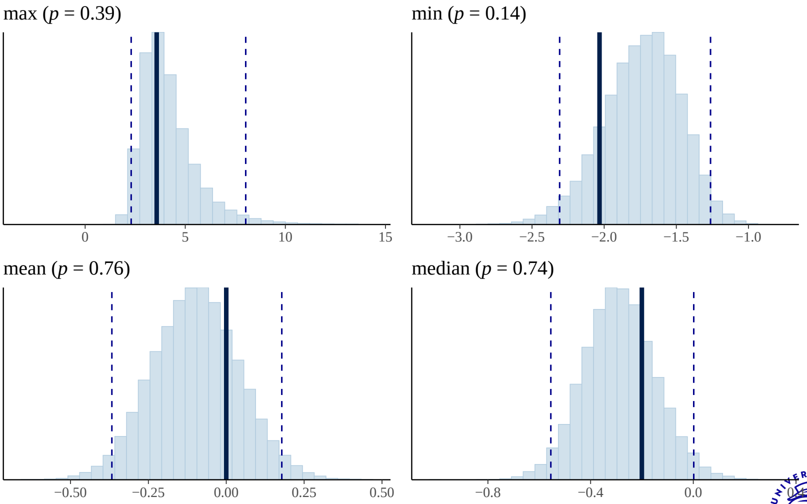


Model evaluation (1/3)

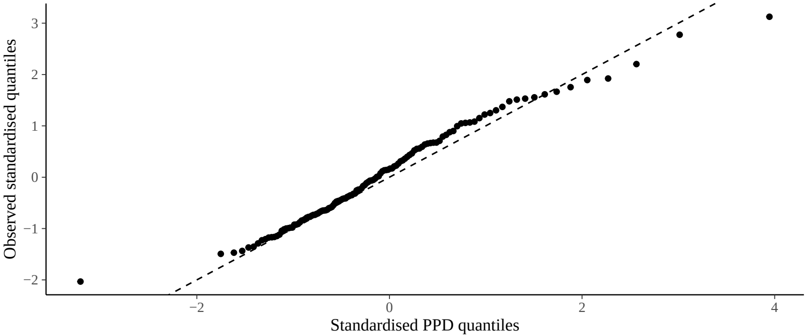




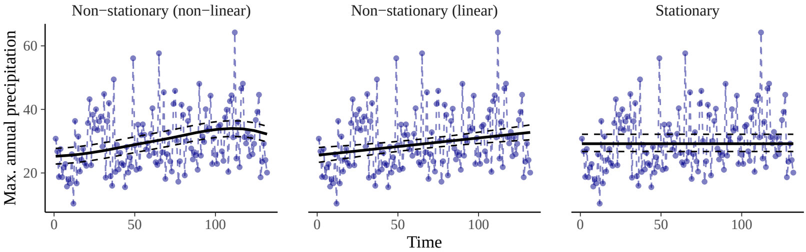
Model evaluation (2/3)



Model evaluation (3/3)



Model comparison



Results and discussion

- Models involving time trend show significant increase in accuracy, suggesting that maximum annual precipitation is evolving with time
- Flexibility of piecewise linear trend superior to simple linear time trend with respect to accuracy
- Rising trend observed in first 100 years which wanes in the remaining 30 years – hard to tell whether this is a continued trend as opposed to instability in extremes.
- Observed trend restricted to location where data are obtained – for a more wholistic picture, need to analyse more catchments and include spatial effects.